

# 確率分割の漸近論への解析的組み合わせ論による接近

間野 修平 数理・推論研究系 准教授

## 1 Pitman確率分割

Bayesian Nonparametrics で用いられる Dirichlet 過程 (Ferguson, 1973) は, 可測空間  $\mathcal{X}$  上の測度  $\mu$  と任意の分割  $\{B_1, \dots, B_k\}$  に対し  $(P(B_1), \dots, P(B_k)) \sim \text{Dirichlet}(\mu(B_1), \dots, \mu(B_k))$  となるランダム測度で,

- $W_i \sim \text{Beta}(1, \theta), \text{iid.}$ ,  
 $P_1 = W_1, P_2 = W_2(1 - W_1), P_3 = W_3(1 - W_1)(1 - W_2), \dots$
- $P(\cdot) = \sum_{i=1}^{\infty} P_i \delta_{Y_i}(\cdot), Y_i \sim \mu(\cdot)/\mu(\mathcal{X}), \text{iid.}$

として得られますが (stick-breaking),  $(P_i)$  を降順に並べた  $(P_{(1)}, P_{(2)}, \dots)$ ,  $\sum_{i=1}^{\infty} P_{(i)} = 1$  を Poisson-Dirichlet 分布 (Kingman, 1982) といいます.  $W_i \sim \text{Beta}(1 - \alpha, \theta + i\alpha)$  により 2 母数に拡張され (Pitman & Yor, 1997),  $0 \leq \alpha < 1$  のときは  $\theta > -\alpha$ ,  $\alpha < 0$  のときは  $\theta = -\alpha m, m = 1, 2, \dots$  です. Poisson-Dirichlet 分布に従う頻度の母集団から異なる種類を抽出するとき, 各種類の大きさがなす分割を Pitman 確率分割といい, 標本抽出のモデルとして広く使われています. 大きさ  $n$  の標本において, 大きさが  $i$  の種類の数を  $s_i$ , 種類の数の和を  $\sum_{i=1}^n s_i = k$ , 降順の大きさを  $l_i$  とします.

例 1  $5 + 2 + 2 + 1$  のとき  $s_5 = s_1 = 1, s_2 = 2, l_1 = 5, l_2 = l_3 = 2, l_4 = 1$ .

定理 1 (Pitman, 1995) Pitman 確率分割は次の pmf で与えられる.

$$p((s_i)) = \frac{(-1)^n [\theta]_{k;\alpha}}{(-\alpha)^k [\theta]_n} n! \prod_{i=1}^n \binom{\alpha}{i}^{s_i} \frac{1}{s_i!}.$$

ただし  $[x]_{i;a} = x(x+a) \cdots (x+(i-1)a)$ .

Pitman 確率分割を得る方法として, 無限個の無限の客が座れる円卓があり,  $n$  名が  $k$  の円卓を占め, 円卓  $i$  に  $s_i$  が座っているとき,

$$\text{空いた円卓に } \frac{\theta + k\alpha}{\theta + n}, \quad \text{円卓 } i \text{ に } \frac{s_i - \alpha}{\theta + n}$$

で座る, という中華料理店過程が良く知られています. 本ポスターは,

- 標本抽出における大きさの極値による検定
- Poisson-Dirichlet 分布の「裾」

などの興味から, 組み合わせ論を母関数の解析に帰着する解析的組み合わせ論, 特に Flajolet & Odlyzko (1990) により導入された特異点解析を用いて, Pitman 確率分割の大きさの順序統計量の分布を導き, その Poisson-Dirichlet 分布への漸近を考察しています. 詳細は arXiv に投稿しています.

## 2 数論との関連

自然数  $n$  の最大, 最小の素因数を  $P^+(n)$  と  $P^-(n)$  とし,

$$\Psi(x, y) = \#\{n \leq x; P^+(n) \leq y\}, \quad \Phi(x, y) = \#\{n \leq x; P^-(n) > y\}$$

とします. 例えば,  $\Psi(10, 2) = \#\{1, 2, 4, 8\} = 4, \Phi(10, 2) = 4$  です.

定理 2 (Dickman, 1930; Buchstab, 1937)  $x \rightarrow \infty$  のとき

$$\Psi(x, y) \sim \rho(u)x, \quad \Phi(x, y) \sim \omega(u) \frac{x}{\log y}, \quad u = \frac{\log x}{\log y}.$$

ここで,  $\rho(u)$  と  $\omega(u)$  はそれぞれ Dickman 関数, Buchstab 関数で,  $u\rho'(u) + \rho(u-1) = 0, u > 1, \rho(u) = 1, 0 \leq u \leq 1, (u\omega(u))' = \omega(u-1), u > 2, u\omega(u) = 1, 1 \leq u \leq 2$  により特徴づけられる.

定理 3  $(\theta, \alpha) = (1, 0)$  ではランダム置換の巡回置換への分解で,

$$P(L_1 < r) \sim \rho(x), P(L_K > r) \sim (nx)^{-1} \omega(x), n, r \rightarrow \infty, r/n \rightarrow x.$$

## 3 順序統計量の分布

一般化階乗係数は  $[\alpha x]_{n;(-1)} = \sum_{k=0}^n C(n, k; \alpha) [x]_{k;(-1)}$  ( $\alpha \neq 0$ ) で定義され,

$$C(n, k; \alpha) = n! \sum_{i=1}^n \prod \binom{\alpha}{i}^{s_i} \frac{1}{s_i!}$$

と表せますが, 和を  $s_{i < r} = 0, s_{i > r} = 0$  に制限したものを定義し, それぞれ  $C_r(n, k; \alpha), C^r(n, k; \alpha)$  で表します. 前者は associated generalized factorial coefficient として知られています. 指数母関数はそれぞれ

$$f_{r,k}(u) := \sum_{n=rk}^{\infty} C_r(n, k; \alpha) \frac{u^n}{n!} = \frac{1}{k!} \left[ (1+u)^\alpha - \sum_{i=0}^{r-1} \binom{\alpha}{i} u^i \right]^k,$$

$$f_k^r(u) := \sum_{n=k}^{rk} C^r(n, k; \alpha) \frac{u^n}{n!} = \frac{1}{k!} \left[ \sum_{i=1}^r \binom{\alpha}{i} u^i \right]^k$$

です. さらに,  $\sum_{i=r+1}^r s_i < t$  で制限したものを  $C^{r(t)}(n, k; \alpha)$  で表すと, 指数母関数は  $f_k^{r(t)}(u) = f_k^r(u) + \sum_{i=1}^{t-1} f_{r+1,i}(u) f_{r,k-i}(u)$  で与えられます.

補題 1 Pitman 確率分割において大きさの順序統計量の周辺分布は

$$P(L_i \leq r) = \sum_{k=i}^{n-r(i-1)} \frac{(-1)^n [\theta]_{k;\alpha}}{(-\alpha)^k [\theta]_n} C^{r(t)}(n, k; \alpha), \quad i = 1, 2, \dots, K.$$

ただし  $i = 1$  では和は  $k = \lceil n/r \rceil, \dots, n$  をとる.

証明  $\alpha$  が既知のとき, 種類の数の和  $K$  は  $\theta$  の十分統計量で

$$P(K = k) = \frac{(-1)^n [\theta]_{k;\alpha}}{(-\alpha)^k [\theta]_n} C(n, k; \alpha)$$

であることから,  $K$  で pmf を条件付けることより従う.  $\square$

## 4 極値の漸近論

$\alpha > 0$  での  $r \asymp n$  における最大値の分布に  $\alpha = 0$  を代入すれば,  $\alpha = 0$  での分布が得られます (Pitman & Yor, 1997). しかし,  $\alpha > 0$  で最小値は 1 に確率収束しますが (Sibuya & Yamato, 2001),  $\alpha = 0$  では 1 とは限りません (Arratia & Tavaré, 1992). このように, 極値の漸近は複雑ですので, (i)  $r \rightarrow \infty, r \asymp n$ , (ii)  $r \rightarrow \infty, r = o(n)$ , (iii)  $r = O(1)$  の 3 通りについて考えました. ここでは  $\alpha > 0$  での最小値について紹介します.

定理 4  $0 < \alpha < 1, \theta > \alpha$  のとき, "Buchstab 関数" は退化する, つまり

$$P(L_K > r) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} n^{-\theta-\alpha}, \quad n, r \rightarrow \infty, r \asymp n.$$

定理 5  $0 < \alpha < 1, \theta > \alpha$  のとき,  $n \rightarrow \infty, r = 2, 3, \dots$  に対し

$$P(L_K \geq r) \sim \frac{\Gamma(1+\theta)}{\Gamma(1-\alpha)} n^{-\theta-\alpha} c_{\alpha,r}^{-1-\frac{\theta}{\alpha}}, \quad c_{\alpha,r} = \sum_{j=1}^{r-1} \binom{\alpha}{j} (-1)^{j+1}.$$

証明 指数母関数に Cauchy-Goursat の定理を適用すると

$$\begin{aligned} P(L_K \geq r) &= (-1)^n \frac{n!}{[\theta]_n} [u^n] \sum_{k=0}^{\infty} \binom{-\frac{\theta}{\alpha}}{k} \left[ (1+u)^\alpha - \sum_{j=0}^{r-1} \binom{\alpha}{j} u^j \right]^k \\ &= \frac{n!}{[\theta]_n 2\pi\sqrt{-1}} \oint \frac{(f_{\alpha,r}(u))^{-\frac{\theta}{\alpha}}}{u^{n+1}} du, \end{aligned}$$

ただし

$$f_{\alpha,r}(u) = (1-u)^\alpha - \sum_{j=1}^{r-1} \binom{\alpha}{j} (-u)^j.$$

Rouché の定理から  $f_{\alpha,r}(u) = 0$  は  $|u| \leq 1$  に根をもたず, 被積分関数の極は原点のみである. 分岐を  $[1, \infty)$  とし, それを  $1/n$  の距離で避ける以外の部分は  $|u| = 1 + \eta, \eta > 0$  を通る積分路をとると積分の値が求まる.  $\square$

